Use Case 3 in Health – Business Cases

Data warehousing in healthcare

KW Partner: FU Berlin

1 Overview

Challenge
A large health insurance company combines its data in a data warehouses to ensure data integration and consistency

Solution
Introduce a common terminology for healthcare data and wrap all legacy data in this terminology

Why a Semantic solution
The provision of a common terminology allows for a data integration and consistency checking solution, in which a computer system can determine the relationship between data items and identify contradictions

Key Business Benefits
There is a cost in time and money involved in a hard-coding of system and data integration, as well as costs to business where integration or consistency is not being handled (i.e. the costs of not having access to data, or having inconsistent data)

Keys components

Existing Software
Data warehousing system

Research and development
Ontology-based integration
Ontology-based proofs
Ontologies for measures
Semantic rules

Technology locks
Data integration
Logical consistency checking
Ontology development/re-use

Companies often administer their data, due to its sheer volume, using data warehousing solutions. A data warehouse is an isolated database which is used across an enterprise to combine data from different data stores and serve all business task-supporting systems with a unified view of the business data. It is characterised by a strict separation of operational and decision-making data and systems.

In this use case, a large health insurance company in Germany uses a Cognos data warehousing solution to administer its data. The business data is being stored in various computer systems, does not share the same data formats and is not integrated with one another. As a result, searching for particular data involves individual searches with the possibility that data is not found because the correct system was not searched. A data warehouse is seen as a solution to this.

Another problem is that data may be changed or updated in one system which may have an affect on data in another system, yet through the lack of any explicit associations between the data, this new inconsistency may not be detected.

As a result, this company has chosen to implement a data warehouse system with a data integration and consistency checking solution. In proposing such a solution, the Semantic Web can be considered.

2 Current Practices and Technologies

2.1 Current business practises

A typical data warehouse architecture consists of a data pool, an archiving system, and a metadata store. The metadata store contains figures, dimensions, reports, cubes and rules. To model a data warehouse system, a modelling tool must take a decentralised approach, offer good visualisation capabilities, be intranet/Internet based, be simple to use, support namespaces and consistency.

Semtation GmbH offers the SemTalk (www.semtalk.com) tool which is an add-on to Microsoft Visio. It offers a decentralised modelling on the basis of a central model, an object-oriented approach, and support for Semantic Web technologies.

An important aspect of the data warehouse modelling is the description of rules. The types of rules that can be described are data reduction, data format simplification, the combination and separation of attribute values, calculations, data correction and consolidation. Dimensions and operational figures can be defined centrally, including rules governing their use. Through the description of such rules in the metadata of the data warehousing system, data integration and consistency checking can be included.

2.2 System requirements Analysis

At present the rules in the data warehousing system model are not very user-friendly, i.e. they use a very unreadable syntax that will not likely be understood by others apart from the rule developer. The introduction of ontologies is a basis for describing the rules in an open standard using a common, shared terminology (in this case, for the healthcare company and its peer companies). This helps make rules understandable, re-usable, and maintainable and ties them to a formal semantic model to support reasoning and consistency. It also forms the basis for more complex search queries and more accurate search results e.g. “How many leg fractures were there in Zehlendorf last year?” will only return a correct result if the system understands all diagnoses which match ‘leg fracture’ and can correctly identify which occurred in the district of Zehlendorf. Open issues in considering this approach (represented in Figure 6.1) are:

- The description of rules based on an ontology
- The definition of measures in ontologies

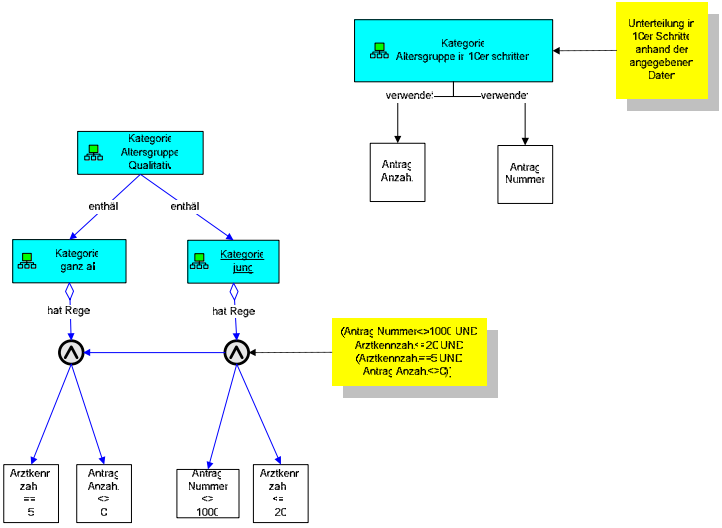

Figure 1 – Measures and rules in the data warehousing system (in German)

2.3 Review of the current systems

There are many data warehouse systems on the market, e.g. SAP Business Information Warehouse, SAS 9.1 and IBM DB2 Business Intelligence.

There are also commercial tools for metadata management in data warehouse systems. This includes Ardent MetaStage, IBM DataGuide, Microsoft Repository, Sybase WCC and Viasoft Rochade.

Current data warehouse environments exhibit no or insufficient support for consistent and comprehensive metadata management [1]. Metadata models are proprietary and typically in a warehouse a number of independent and heterogeneous repositories will co-exist.

In fact, metadata solutions are a key need of data warehousing [2], above all those which offer business process modelling and the integration of the heterogeneity from the operational source to the user interface.

References

1. Do, Hong Hai; Rahm, Erhard. "On Metadata Interoperability in Data Warehouses", Technical Report 1-2000, Institute for Computer Science, University of Leipzig, 2000.
Do, Hong Hai; Stöhr, Thomas; Rahm, Erhard; Müller, Robert; Dern, Gernot. „Evaluierung von Data Warehouse-Werkzeugen", Proc. Data Warehousing (DW) 2000, Friedrichshafen, Nov. 2000

